

## NEWS AND VIEWS

## OPINION

## Purging putative siblings from population genetic data sets: a cautionary view

ROBIN S. WAPLES\*  and  
ERIC C. ANDERSON† 

\*NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112, USA; †NOAA Fisheries, Southwest Fisheries Science Center, 110 McAllister Way, Santa Cruz, CA 95060, USA

Interest has surged recently in removing siblings from population genetic data sets before conducting downstream analyses. However, even if the pedigree is inferred correctly, this has the potential to do more harm than good. We used computer simulations and empirical samples of coho salmon to evaluate strategies for adjusting samples to account for family structure. We compared performance in full samples and sibling-reduced samples of estimators of allele frequency ( $\hat{P}$ ), population differentiation ( $\hat{F}_{ST}$ ) and effective population size ( $\hat{N}_e$ ). Results: (i) unless simulated samples included large family groups together with a component of unrelated individuals, removing siblings generally reduced precision of  $\hat{P}$  and  $\hat{F}_{ST}$ ; (ii)  $\hat{N}_e$  based on the linkage disequilibrium method was largely unbiased using full random samples but became increasingly upwardly biased under aggressive purging of siblings. Under nonrandom sampling (some families over-represented),  $\hat{N}_e$  using full samples was downwardly biased; removing just the right ‘Goldilocks’ fraction of siblings could produce an unbiased estimate, but this sweet spot varied widely among scenarios; (iii) weighting individuals based on the inferred pedigree (to produce a best linear unbiased estimator, BLUE) maximized precision of  $\hat{P}$  when the inferred pedigree was correct but performed poorly when the pedigree was wrong; (iv) a variant of sibling removal that leaves intact small sibling groups appears to be more robust to errors in inferences about family structure. Our results illustrate the complex challenges posed by presence of family structure, suggest that no single optimal solution exists and argue for caution in adjusting population genetic data sets for the presence of putative siblings without fully understanding the consequences.

**Keywords:** allele frequency, effective population size, family structure, genetic differentiation, precision

Received 19 October 2016; revision received 10 January 2017; accepted 12 January 2017

### Introduction

Random sampling is a convenient theoretical and statistical construct. It is easy to implement in computer simulations but difficult or impossible to achieve in the real world. A truly random sample has two important features: *equal opportunity*—every individual in the focal population must have the same probability of appearing in the sample; and *independence*—the probability that an individual will be sampled does not depend on whether any other individuals are in the sample. A violation of the second criterion occurs when relatives are likely to be sampled together. Recent attention related to nonrandom sampling has focused on three key points:

- 1 Nonrandom samples, with some families being over-represented compared to their occurrence in the population, are common in natural populations, particularly for species with relatively high fecundity that are sampled at early life stages (Hansen *et al.* 1997; Goldberg & Waits 2010; Whiteley *et al.* 2012);
- 2 Presence of family structure in samples can affect some population genetic analyses (Anderson & Dunham 2008; Rodríguez-Ramilo & Wang 2012);
- 3 Recent advances in DNA technology and analytical methodology have improved the ability to identify relatives in natural populations (Ashley *et al.* 2009; Jones & Wang 2010; Almudevar & Anderson 2012).

Based on these three observations, many researchers (e.g. Hess *et al.* 2015) now routinely remove all but one member of a putative sibling group before using population genetic data in downstream analyses—an approach that is now is often regarded as ‘best practices’ (Peterman *et al.* 2016). At first glance, the premises seem sound and the conclusion seems logical—what could go wrong?

Actually, a lot can go wrong. At least three major problems can arise when one attempts to purge putative siblings from samples. (i) Siblings occur naturally in all populations at frequencies that are inversely related to effective population size ( $N_e$ )—and this fact forms the basis of the sibship method for estimating  $N_e$  (Wang 2009). Indiscriminant removal of putative siblings thus risks erasing part of the evolutionary signal of small populations, making them look more like large or infinitely large ones. (ii) Even if sibling removal is effective in reducing the appearance of nonrandom sampling, it also reduces the sample

Correspondence: Robin S. Waples, Fax: +1 206 860 3335; E-mail: robin.waples@noaa.gov

size, which sets up an inevitable trade-off with respect to precision and statistical power that should be formally considered. (iii) Methods for sibling inference are not infallible, particularly for identification of half-siblings (HS) or other more distant relatives. Results of sibship reconstruction often differ depending on the method used (Ringler *et al.* 2015) and, for a given method, can differ depending on the type of markers used (Linlökken *et al.* 2016). Therefore, researchers interested in pursuing removal of putative siblings should consider the consequences of imperfect ability to identify relatives.

As far as we can determine, the only analyses for which sibling removal has been convincingly demonstrated to improve performance are Bayesian clustering methods such as STRUCTURE (Pritchard *et al.* 2000), for which family groups can be mistaken for separate ‘populations’ (Anderson & Dunham 2008; Rodríguez-Ramilo & Wang 2012). Whether this reflects a shortcoming of the method itself or rather a problem of interpretation is an open question. Rodríguez-Ramilo *et al.* (2014) found that some other clustering methods that do not make assumptions about Hardy–Weinberg equilibrium and linkage equilibrium were much less sensitive to family groups than STRUCTURE. This suggests that the effects of siblings are case-specific and depend on the particular analyses involved.

Because related individuals in a sample create a mismatch between the variance of estimated allele (or genotype or gamete) frequencies and the variance predicted under the assumption of random, independent sampling, it is clear that presence of siblings can bias results of statistical tests that compare the observed deviations in statistics (like allele frequencies) with their expectations under random sampling theory. Accordingly, researchers should remain vigilant about effects of siblings on statistical hypothesis tests. However, this does not necessarily mean that eliminating putative siblings from samples is a sound strategy in general. For example, although sibling removal has been advocated to improve point estimates of allele frequency (Goldberg and Waits 2010), it can actually reduce performance of the estimator. This potential downside to sibling removal indicates that it is risky to routinely remove putative siblings from genetic data sets without a clear understanding of the consequences.

Here, we use computer simulations to evaluate the practical consequences of removing siblings from both random and nonrandom samples, generated under several mating models. We focus on two single-locus metrics (allele frequency and  $F_{ST}$ ) and one-two-locus metric [linkage disequilibrium (LD) between pairs of loci, which is often used to estimate  $N_e$ ]. Removing putative siblings before estimating  $N_e$  is less common than for other downstream analyses, but it does occur (e.g. Peterman *et al.* 2016). Results indicate that even if family relationships are known without error, removing siblings can degrade precision of estimates of allele frequency and  $F_{ST}$  and bias estimates of  $N_e$ . The results argue for caution by researchers who are tempted to try purging siblings from samples from real populations.

To better understand these simulation results, we next undertake an analytical treatment of the effect of sibling removal on allele frequency estimation. We consider sibling removal as a special (extreme) case of unequal weighting of the information from each individual in a sample. Use of such weighting schemes to obtain the best linear unbiased estimator (BLUE) for a parameter is well known in statistics, and we review McPeck *et al.* (2004)’s application of the BLUE approach for estimating allele frequency in the presence of related individuals. Using the same mathematical machinery, we obtain an algorithm to find the optimal sibling removal scheme. Finally, we assess performance of the BLUE approach and several strategies for sibling removal using multiple collections of highly related, juvenile coho salmon (*Oncorhynchus kisutch*). Our results indicate that, if the true family relationships are known without error, the BLUE approach produces an estimate with lower variance than is obtained either by any variation of sibling removal or use of the full (unweighted) sample; however, using the full sample is better than BLUE when sibling identification is not reliable. A method of partial sibling removal that leaves intact family groups no larger than two performed nearly as well as the naïve estimator when inferred pedigrees were unreliable and generally better than the naïve estimator when the inferred pedigree was correct.

## Materials and methods

### Computer simulations

*Reproduction and sampling.* The simulations had the following features. Modelled populations were closed to immigration and emigration, had discrete generations, a constant number ( $N$ ) of mature individuals and separate sexes with an equal number ( $N/2$ ) of males and females. Selfing was not allowed. After initialization, each population was allowed to reproduce for 10 generations, following which a sample of  $S$  individuals for genetic analysis was produced by the  $N$  parents in generation 10 (Table 1). Genotypes at 100 diallelic (SNP) loci were tracked in each individual. For each parameter set, 100 replicate samples of  $S$  offspring were generated as described in more detail in Supporting information.

Three mating systems were considered (Table 1). In the monogamy and random models, reproduction in generations 1–10 was by Wright–Fisher ideal populations with and without monogamy, respectively (so  $N_e \approx N$ ). In the final generation, samples of  $S$  progeny were produced in two ways: (i) using the same two ideal mating systems (random samples), with the only difference being that  $S$  could be larger than  $N$ ; or (ii) by allowing some pairs of parents to produce large families (to mimic nonrandom, family-correlated samples). In the latter case, the effective number of parents that produced the sample ( $N_b$ ) was less than  $N_e$ . In the third (mixed) model of reproduction, in

**Table 1** Summary of the processes of reproduction and sampling in the three simulation models. In all models, the population was allowed to reproduce at constant size  $N$  for 10 generations, and then samples were produced by the parents in generation 10

Model	Reproduction in generations 1–10	Sampling*
Random	Wright–Fisher without selfing Constant $N$ ; $N_e = N + 0.5 + 1/2N$	F1, F3, F9
Monogamy	Wright–Fisher with monogamy Constant $N$ ; $N_e = N + 0.5 + 1/2N$	F1, F3, F9
Mixed	Overdispersed variance in reproductive success Maximum family size/reproductive event = 9 or 30 Constant $N$ ; $N_e < N$	Random; same scheme as for reproduction†

\*F1, F3, F9 are maximum family sizes for each pair of randomly chosen parents. F1 produces a random sample. For F3 and F9, family size at each reproductive event is drawn from an even distribution 2–3 or 2–9, respectively, and this produces samples with excesses of siblings.

†In the mixed model, samples were randomly produced by the same process used for reproduction in generations 1–10, with the exception that sample size could exceed  $N$ .

generations 1–10, some pairs of parents were allowed to produce  $>1$  offspring per mating episode, so variance in reproductive success was greater than random, leading to  $N_e < N$  each generation. Only random samples were taken in the final generation for the mixed model, but these also were drawn from parents for which  $N_e < N$ . Full details of these mating models and sampling of juvenile offspring can be found in Supporting information.

**Sibling removal.** In addition to considering the full samples, we evaluated subsets of the samples from which some or all siblings were removed. Removal was based on the recorded pedigree and did not consider potential errors in sibling identification. In fractional removals, individuals from the full sample were evaluated one at a time to determine whether they should be included in the reduced sample. If the individual was a relative of any other individual already in the reduced sample, it was excluded with probability  $\beta$ . Values evaluated were  $\beta = 0.25, 0.5, 0.75, 0.9, 0.95$  and 1. With  $\beta = 1$ , all but one member of each family was excluded; we refer to this as 100% sibling removal. In the random and mixed mating models, sibship exclusion was considered two ways: excluding all siblings or only full-siblings (FS).

All simulations and data analyses were conducted in R 3.2.3 (RCT 2015). The 25 different scenarios that were simulated are summarized in Table 2.

#### Sibling production

The random mating and mixed models produced three classes of offspring based on their one-generation pedigree: FS, HS and unrelated (U). Only FS and U offspring were produced in the monogamy model. The proportions of siblings produced are expected to be simple functions of  $N_e$  and the mating model. This relationship, based on a simplification of eqn 10 in Wang (2009) that ignores the (generally small) correction for departures from Hardy–Weinberg equilibrium, is as follows:

$$\frac{1}{N_e} = \frac{Q_{HS} + 2Q_{FS}}{4}, \quad (\text{eqn 1})$$

where  $Q_{HS}$  is the fraction of pairs that are HS (maternal and paternal HS combined) and  $Q_{FS}$  is the fraction that are FS. It can be shown (Ackerman *et al.* 2017) that eqn 1 yields the same estimate of  $N_e$  as the parentage-analysis-without-parents approach of Waples & Waples (2011), which calculates inbreeding  $N_e$  based on the vector of numbers of offspring produced by each parent. Simple rearrangement of eqn 1 produces the expected frequencies of sibships:

$$Q_{HS} + 2Q_{FS} = \frac{4}{N_e}. \quad (\text{eqn 2})$$

Observed fractions of siblings produced in the simulations were tracked and compared with the expected fractions based on eqn 2, and realized  $N_e$  was calculated from the pedigree using the following equation from Waples & Waples (2011):

$$N_e = \frac{2S - 1}{\frac{\sum(k_i^2)}{2S} - 1}, \quad (\text{eqn 3})$$

where  $S$  is the number of individuals in the sample and  $k_i$  is the number of sampled offspring produced by the  $i$ th parent. Realized effective size was calculated separately for each sex ( $N_{eM}$ ,  $N_{eF}$ ), and the overall  $N_e$  was calculated as  $N_e = 4N_{eM}N_{eF}/(N_{eM} + N_{eF})$  (Wright 1938).

#### Data analysis

Analyses of simulated data focused on three metrics.

**Allele frequency.** For each locus, true allele frequency ( $TrueP$ ) was calculated as the mean frequency in the  $N$  parents from generation 10. For each replicate sample of offspring, estimated allele frequency ( $\hat{P}$ ) for each locus was

**Table 2** Scenarios considered in the simulations

Scenario	$N$	$N_e$	$S$	Mating	Sampling	Sibling removal*
A	20	20.5	100	Monogamy	Random	FS
B	20	20.5	100	Monogamy	F9 <sup>†</sup>	FS
C	40	40.5	40	Monogamy	Random	FS
D	40	40.5	40	Monogamy	F9	FS
E	40	40.5	100	Monogamy	F9	FS
F	40	40.5	100	Monogamy	F3 <sup>‡</sup>	FS
G	40	40.5	100	Monogamy	Random	FS
H	100	100.5	50	Monogamy	Random	FS
I	250	250.5	50	Monogamy	Random	FS
J	500	500.5	40	Monogamy	Random	FS
K	500	500.5	40	Monogamy	F9	FS
L	500	500.5	100	Monogamy	Random	FS
M	40	21.0 <sup>§</sup>	100	MixedF9 <sup>¶</sup>	Random**	FS + HS
N	40	21.3 <sup>§</sup>	100	MixedF9 <sup>¶</sup>	Random**	FS
O	100	15.3 <sup>§</sup>	40	MixedF9 <sup>¶</sup>	Random**	FS + HS
P	100	15.7 <sup>§</sup>	40	MixedF9 <sup>¶</sup>	Random**	FS
Q	500	16.5 <sup>§</sup>	40	MixedF9 <sup>¶</sup>	Random**	FS
R	500	8.2 <sup>§</sup>	40	MixedF30 <sup>††</sup>	Random**	FS + HS
S	40	40.5	100	Random	Random	FS + HS
T	40	40.5	100	Random	Random	FS
U	40	40.5	100	Random	F9	FS + HS
V	40	40.5	100	Random	F9	FS
W	100	100.5	40	Random	F9	FS + HS
X	100	100.5	40	Random	F9	FS
Y	500	500.5	40	Random	Random	FS

FS, full-siblings; HS, half-siblings.

\*No HS are produced with monogamy.

<sup>†</sup>Maximum family size = 9 siblings.

<sup>‡</sup>Maximum family size = 3 siblings.

<sup>§</sup> $N_e < N$  in the mixed mating model because larger families were produced than with random mating.

<sup>¶</sup>Parameters:  $\alpha$  (probability of each pair producing >1 offspring) = 0.5; maximum family size = 9 siblings.

\*\*Sampling was random but the mixed mating model produced large families.

<sup>††</sup>Parameters:  $\alpha = 0.1$ ; maximum family size = 30 siblings.

calculated as the mean across all individuals in the sample. For each locus, the root-mean-squared-error (RMSE) of allele frequency was calculated as

$$\text{RMSE}_P = \sqrt{\Sigma(\hat{P} - \text{True}P)^2 / 100},$$

with the summation taken across all 100 replicate samples. An overall RMSE was then calculated as the mean  $\text{RMSE}_P$  across all 100 loci. The consequences of sibling removal for allele frequency estimation were quantified as the relative  $\text{RMSE}_P$  ( $\phi_P$ ), which is the ratio of  $\text{RMSE}_P$  for the subsample and the full sample. Values of  $\phi < 1$  indicate that sibling removal increased precision of  $\hat{P}$ , while values  $>1$  indicate that sibling removal degraded performance.

We can also define an effective sample size in a manner analogous to the definition of effective population size. Let an 'ideal' sample be one in which individuals are drawn randomly and every individual provides completely independent information. For estimation of

parental allele frequency using offspring, an ideal sample would be one containing  $S$  unrelated individuals, in which case the variance of  $\hat{P}$  is just the binomial sampling variance:

$$\text{Var}(\hat{P}) = E(\text{True}P - \hat{P})^2 = \text{True}P(1 - \text{True}P)/(2S). \quad (\text{eqn } 4)$$

In this ideal sample, the effective sample size is the same as the sample size ( $\text{ESS}_{\text{ideal}} = S$ ). For a nonideal sample, ESS is defined as the size of an ideal sample that would be expected to produce the observed value of  $\text{Var}(\hat{P})$ . Substituting mean-squared-error  $\text{MSE}_P$  for  $E(\text{True}P - \hat{P})^2$  and rearrangement of the above produces

$$\text{ESS} = \text{True}P(1 - \text{True}P)/[2\text{MSE}_P]. \quad (\text{eqn } 5)$$

Equation 5 was used to calculate effective sample sizes for the original samples ( $\text{ESS}_{\text{full}}$ ) and the sibling-reduced samples ( $\text{ESS}_{\text{reduced}}$ ). ESS was calculated for each locus, and the harmonic mean across all 100 loci was used as the overall ESS.

*Population differentiation.* For each parameter set, the 100 replicate daughter populations were divided into 50 pairs, and Nei's (1973)  $F_{ST}$  was calculated for each locus as

$$F_{ST} = \frac{H_T - H_S}{H_T},$$

where  $H_S$  is the average expected heterozygosity within the two populations and  $H_T$  is the total expected heterozygosity based on mean allele frequencies across the populations. For each pair of populations, an overall mean  $F_{ST}$  ( $\bar{F}_{ST}$ ) was calculated across all 100 loci using data for all  $N$  parents, and this was considered the parametric (true)  $F_{ST}$ . Unbiased estimates of  $F_{ST}$  ( $\hat{F}_{ST}$ ) from samples of progeny were adjusted for sampling error by subtracting the quantity  $1/(2S)$  (Wright 1978; Chakraborty & Leimar 1987). Loci monomorphic in both samples were excluded. Across the 50 pairs of populations, and for each sibling-reduced sample size, RMSE of  $F_{ST}$  was calculated as

$$RMSE_{F_{ST}} = \sqrt{\Sigma(\bar{F}_{ST} - \hat{F}_{ST})^2 / 50},$$

and the relative RMSE of  $F_{ST}$  ( $\phi_{F_{ST}}$ ) was calculated analogously to  $\phi_p$  described above.

*Effective population size.* True effective size under the monogamy and random mating models was  $N_e = N + 0.5 + 1/(2N)$  (Balloux 2004); these values are shown in Table 2, but for simplicity in the text only the whole number is used. In the mixed mating model, the possibility that a parental pair would be allowed to produce multiple offspring at each draw led to overdispersed variance in reproductive success and  $N_e < N$ . Realized  $N_e$  in the mixed model was calculated using eqn 3, which was also used to track the realized effective number of parents that produced the sibling-reduced samples for all mating models.

For each sample,  $N_e$  was estimated using the LD method (Waples & Do 2008), and an overall harmonic mean  $\hat{N}_e$  was calculated across all replicates. Alleles with frequencies  $< 0.05$  were excluded. In addition, RMSE of  $1/\hat{N}_e$  (as recommended by Wang 2009) was calculated as described above for  $\hat{P}$ , and results were used to calculate a relative RMSE ( $\phi_{N_e}$ ) of estimates of effective size.

#### *Analytical calculation of effective sample size in empirical samples*

Estimating allele frequency by removing siblings can be thought of as giving a weight of '1' to all unrelated individuals in the sample and a weight of '0' to all but one sibling from each family. A more general approach would allow noninteger weights for each individual, based on information about their degree of relatedness to others in the sample. McPeck *et al.* (2004) used this approach to derive a best linear unbiased estimator (BLUE) for allele frequencies when the sample consists of individuals of

known relationship. Some programs for parentage analysis (e.g. Colony; Wang & Santure 2009) also use the inferred pedigree to update estimates of allele frequency. We use the same mathematical analysis framework to compute the effective sample size of any weighted or sibling-removed sample, conditional on the pedigree connecting the sample members. To allow comparison of weighted and sibling removal methods for estimating allele frequency, we also developed a greedy algorithm to identify the optimal sibling elimination scheme, given the known family relationships among individuals in the sample (See Supporting information for details).

We evaluated performance of the estimators using 70 collections of juvenile coho salmon taken from streams in California and Southern Oregon (described in Gilbert-Horvath *et al.* 2016). Each collection, which varied in size from 10 to 150 individuals (mean = 59), was genotyped at 95 SNPs and analysed with Colony (Version 2, Wang 2004; Wang & Santure 2009) to infer FS and HS. We then used the inferred pedigree to compute the effective sample size under five different weighting schemes: (i) naïve: full sample with no weighting; (ii) BLUE: weighting according to the best unbiased linear estimator; (iii) optimal-z: the minimum variance sibling removal scheme found using our greedy algorithm; (iv) Yank-1: HS were ignored, and individuals were reduced randomly from full-sibships of size  $\geq 3$  until only one member of the family remained; (v) Yank-2: identical to Yank-1 except individuals were reduced randomly from FS groups of size  $\geq 3$  until only two members of the family remained. The Yank-1 scheme was used in Garza *et al.* (2014) and Gilbert-Horvath *et al.* (2016) for investigations into juvenile steelhead (anadromous *Oncorhynchus mykiss*) and coho salmon.

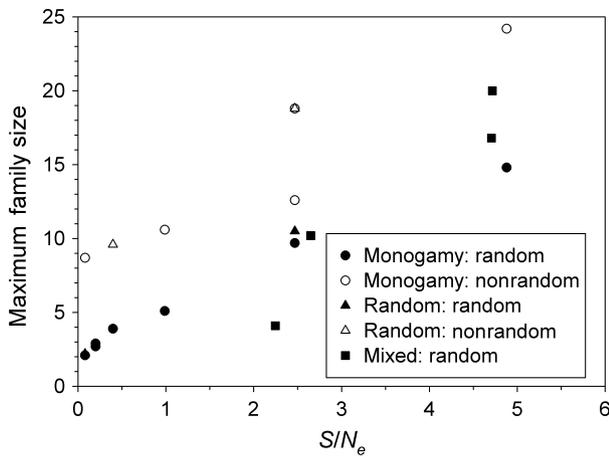
In the empirical evaluations under the 'Sample Related' scenario, we assumed that the inferred pedigree was correct and used it to compute the effective sample size for all five methods. To evaluate consequences of errors in sibling reconstruction, in the 'Sample Unrelated' scenario the gene copies carried by each member of each collection were permuted within loci (holding missing data positions constant), yielding for each collection a permuted sample in which there was no longer any relationship among the sample members. These permuted samples were analysed with Colony 2, weighting schemes were calculated using the Colony-inferred pedigree, and effective sample sizes for all schemes were calculated using the fact that all scrambled sample members are effectively unrelated. This provides an indication of the performance of sibling elimination under pedigree reconstruction errors.

## Results

Results are summarized in Figs 1–8 and S1–S6; see Tables S1 and S2 (Supporting information) for more details.

#### *Sibling frequency and family size*

Observed frequency of siblings in the samples agreed closely with those expected from eqn 2 (Fig. S1, Supporting



**Fig. 1** Relationship between mean maximum family size in the full sample and the ratio of sample size to effective population size ( $S/N_e$ ). Results are for simulated data using three different mating models and random and nonrandom samples. Only random sampling was used with the mixed mating model. Sample size was allowed to be larger than  $N_e$ , as often occurs in sampling early life stages of species with high fecundity. High ratios of  $S/N_e$  can produce large family sizes even in random samples.

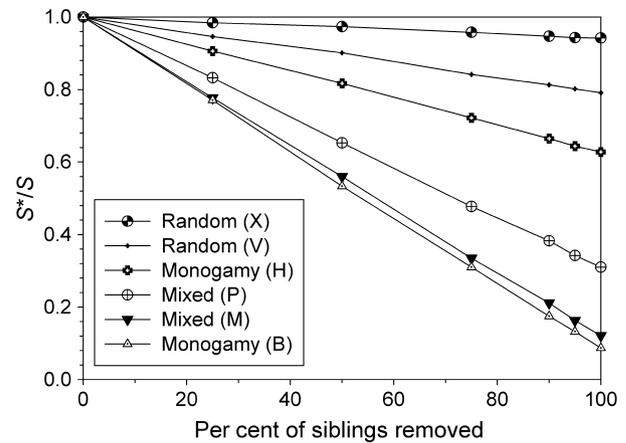
information). Although expected overall frequencies of siblings depend only on realized  $N_e$ , the mix of FS and HS depends on the mating system and the type of samples. Furthermore, the distribution of family sizes in the sample depends strongly on the sample size. Both random and family-correlated samples showed a strong, positive relationship between maximum family size and the ratio  $S/N_e$  (Fig. 1). For the same  $S/N_e$ , random samples had consistently smaller maximum family sizes. However, a random sample for a scenario in which sample size is large compared to effective size can have larger family sizes than are found in some nonrandom samples. Frequencies of FS and HS in the complete samples for the three mating models are shown in Fig. S2 (Supporting information).

#### Reduced sample sizes

Complete sibling exclusion reduced the final sample size by up to 90% or more when  $S$  was larger than  $N_e$  and all siblings were excluded (e.g. scenarios B and M; Fig. 2). At the other extreme, in some scenarios with very small  $S/N_e$  and/or exclusion only of FS (e.g. Scenario X; Fig. 2), the final sample sizes were reduced by <10%. Partial removal of siblings had predictably intermediate consequences for sample size.

#### Allele frequency

Removing siblings from random samples of progeny generally reduced precision, such that  $RMSE_p$  was higher after sibling removal than it was for the full data set (Fig. 3, top). This effect was nonlinear, with  $RMSE_p$  rising faster for high levels of sibling removal. With 100% sibling removal,  $RMSE_p$  could be over twice as large as for the



**Fig. 2** Reduced sample size after sibling removal ( $S^*$ ) as a fraction of the full sample size ( $S$ ) in simulated samples produced under three mating models. Letters in parentheses indicate specific scenarios described in Table 2. In some scenarios, aggressive purging of siblings reduces sample size to a small fraction of the original.

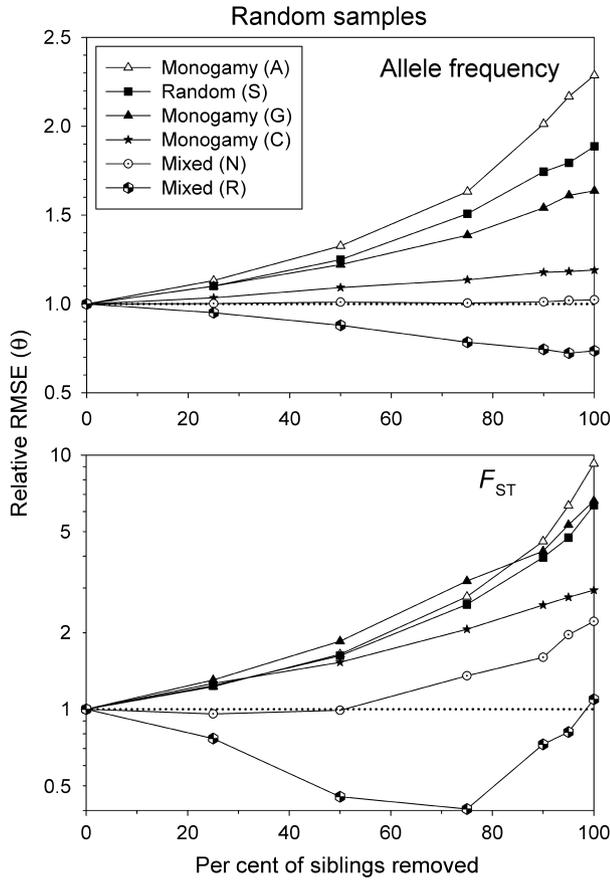
unpurged data set. In some scenarios,  $RMSE_p$  increased only very slightly with removal of siblings. This occurred when few siblings were produced in the first place (Scenario J) or most siblings were HS but only FS were removed (Scenario T) (Table 2), but also in some mixed mating scenarios (e.g. Scenario N, Fig. 3) where sampling was random but many siblings were produced. In one extreme scenario using the mixed mating model (Scenario R), where 10% of the randomly chosen pairs of parents were allowed to produce up to 30 FS,  $RMSE_p$  of  $\hat{P}$  dropped by 30% with removal of all siblings (Fig. 3, top).

Surprisingly, very similar patterns were found in most scenarios with nonrandom samples: removing siblings increased  $RMSE_p$  (by up to ~50%) and reduced precision compared to the full samples (Fig. 4, top). In one case (Scenario V, random mating with family-correlated sampling that primarily produced HS), removing 100% of the FS actually reduced  $RMSE_p$  by 3%.

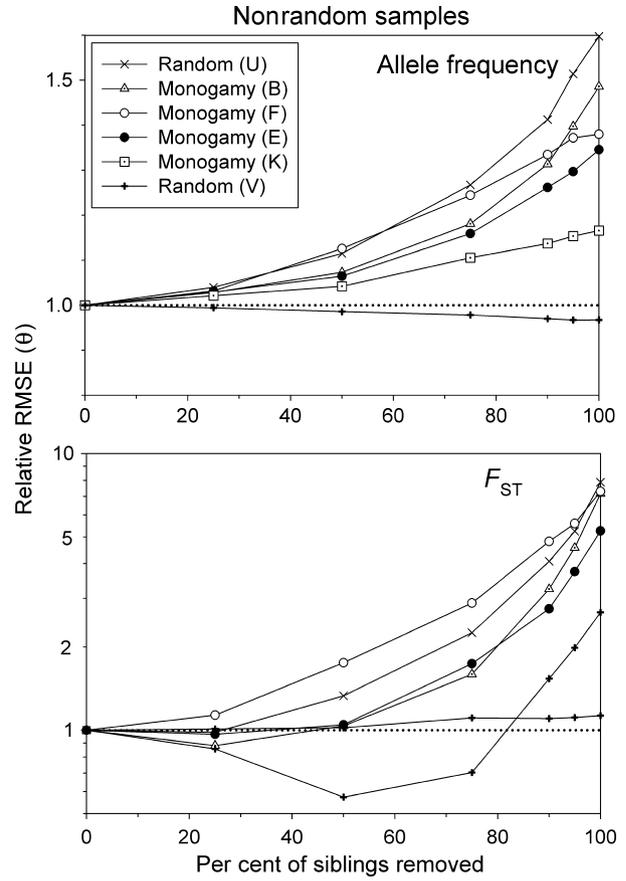
In every scenario considered, regardless whether sampling was random or not, the ratio of effective size of the reduced sample ( $ESS_{\text{reduced}}$ ) to the sibling-reduced sample size ( $S^*$ ) increased with more aggressive purging of siblings (Fig. 5, top). A positive relationship between  $ESS_{\text{reduced}}/S^*$  and the fraction of siblings removed indicates that removing siblings did not increase the variance in  $\hat{P}$  as much as would be expected if all the individuals removed provided completely independent information about allele frequency. However, with the exception of Scenario R, the ratio  $ESS_{\text{reduced}}/ESS_{\text{full}}$  was <1, indicating that effective sample size after sibling removal was less than the original sample size (Fig. 5, bottom).

#### Population differentiation

Results for estimation of  $F_{ST}$  largely paralleled those for estimation of allele frequency: in most scenarios with both



**Fig. 3** Relative RMSE ( $\theta$ ) of allele frequency and genetic differentiation as a function of the percentage of siblings removed from samples simulated under three mating models. Top:  $\theta_p$ . Bottom:  $\theta_{F_{ST}}$  (note the log scale for  $y$ -axis). Values of  $\theta > 1$  indicate that sibling removal reduced precision.



**Fig. 4** As in Fig. 3, but for nonrandom samples. Note the log scale for  $y$ -axis in the bottom panel. Even when sampling is not random and includes excess family structure, purging of siblings often reduces precision for estimating allele frequency and population differentiation.

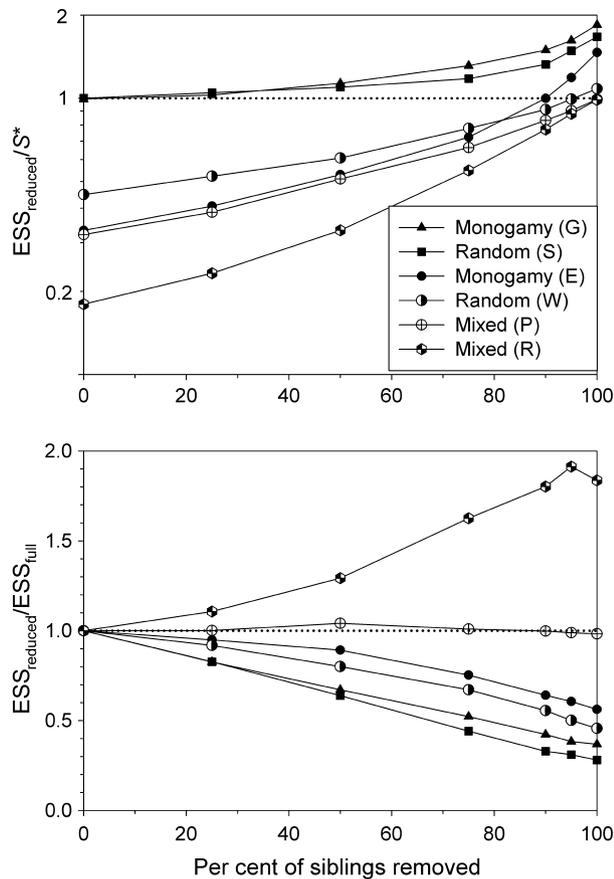
random and nonrandom samples, removing siblings increased  $RMSE_{F_{ST}}$  (Figs 3 and 4, bottom). Two differences in results for  $\hat{P}$  and  $\hat{F}_{ST}$  are worth noting, however. First, the effects were magnified for  $F_{ST}$ : purging all siblings led to almost a 10-fold increase in  $RMSE_{F_{ST}}$  for some scenarios with both random and nonrandom sampling. Second, in some scenarios using the mixed or monogamy mating models,  $RMSE_{F_{ST}}$  was reduced when intermediate levels of siblings were removed. The best result occurred for Scenario R noted above, where  $RMSE_{F_{ST}}$  was reduced by 55–60% when 50–75% of siblings were removed Fig. 4, bottom). However, in this scenario, precision of  $\hat{F}_{ST}$  declined compared to that of the full sample if all siblings were removed.

#### Effective population size

Harmonic mean  $\hat{N}_e$  calculated using eqn 3 closely tracked theoretical true  $N_e$  in scenarios with random sampling and random mating or monogamy (Fig. S3, Supporting information). In the scenarios with nonrandom sampling and in

the mixed mating model (where sampling was random but mating was not), eqn 3 estimated realized  $N_e$  based on the pedigree for the samples, and  $\hat{N}_e$  from LDNe closely tracked the realized  $N_e$  (Fig. S4, Supporting information). When sibling removal was less than about 80%,  $\hat{N}_e$  from LDNe was generally a few per cent higher than realized  $N_e$  from eqn 3. This reflects a slight upward bias in the LDNe version of the LD method, which has been documented elsewhere (e.g. Waples & Do 2010). With 100% sibling removal, the point estimate from eqn 3 is infinity (see eqn 2). Because sibling removal was probabilistic, on occasion all siblings were removed from samples by chance when the probability of removal was high but  $<1$ . Infinite estimates (recorded here as 99 999 for each sex) increased harmonic mean realized  $N_e$ , such that it generally exceeded that from LDNe under very aggressive purging of siblings (Fig. S4, Supporting information).

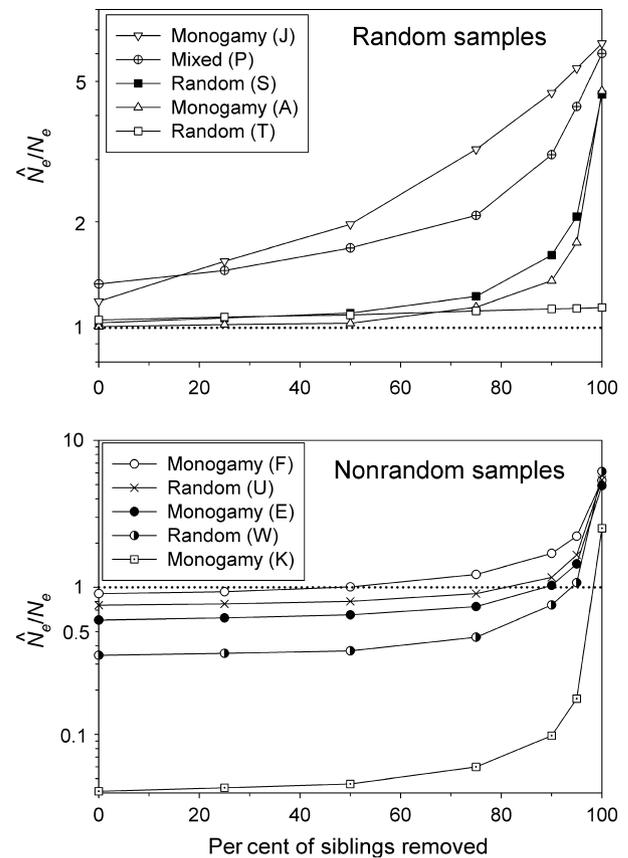
Removing siblings had the expected effect of increasing  $\hat{N}_e$  because truncating largest family sizes reduced disparities in reproductive success among parents. With random sampling,  $\hat{N}_e$  from LDNe was essentially unbiased using the full sample but became increasingly upwardly biased



**Fig. 5** Top: the ratio of effective sample size (ESS) to reduced sample size after removal of siblings ( $S^*$ ). For a given scenario, a positive slope indicates that  $ESS_{\text{reduced}}$  does not decline as fast as  $S^*$  when larger fractions of siblings are removed. Note the log scale on the y-axis. Bottom: the ratio of  $ESS_{\text{reduced}}$  to ESS for the full sample size ( $S$ ). Values >1 indicate that removing siblings increased precision of  $\hat{P}$ .

as a larger fraction of siblings was removed (Fig. 6, top). For example, Scenario J (with a large  $N_e = 500$  and a small  $S = 40$ ) did not produce many siblings, but removing the few siblings that did occur by chance produced an estimate that was over six times true  $N_e$ .

When sampling was nonrandom, estimates of  $N_e$  based on full samples were all downwardly biased, and  $\hat{N}_e$  increased with more aggressive purging of siblings (Fig. 6, bottom). With nonrandom sampling, a transition from underestimation to overestimation of  $N_e$  occurs somewhere along the continuum of fractional sibling removal. This means that (in theory at least) an unbiased estimate could be obtained from nonrandom samples by removing exactly the right fraction of siblings. This ‘sweet spot’, however, varied widely among scenarios. In Scenario F, where the sample was only moderately nonrandom, removal of half of the siblings produced an unbiased  $\hat{N}_e$  (Fig. 6, bottom). But for scenarios U, E, W and K,  $\hat{N}_e$  did not become unbiased until about 80%, 90%, 95% and >95% of siblings were removed, respectively.

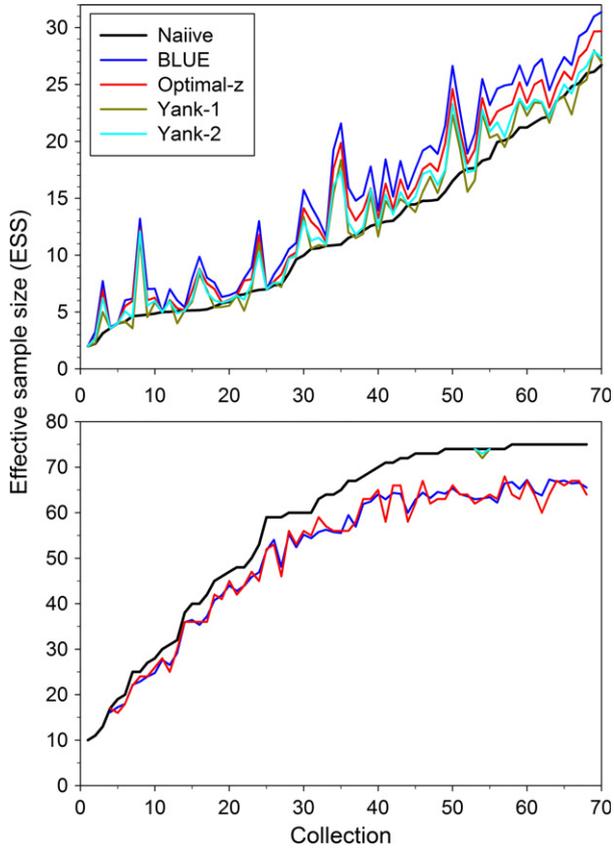


**Fig. 6** Effects of sibling removal on the ratio of  $\hat{N}_e$  (estimated using LDNe; Waples & Do 2008) to true  $N_e$ . Top: random samples. Bottom: nonrandom samples. Note the log scale for the y-axis in both panels. When sampling is random, removing siblings upwardly biases  $\hat{N}_e$ . Purging some siblings from nonrandom samples might reduce downward bias in  $\hat{N}_e$ , but removing too many siblings can create severe upward bias.

Removing siblings from random samples always sharply increased relative  $RMSE_{N_e}$ , except in scenarios where there were few siblings to remove (Fig. S5, top, Supporting information). With nonrandom samples, removing siblings generally lowered  $RMSE_{N_e}$  until sibling removal exceeded the sweet spot where the estimates of  $N_e$  became largely unbiased, after which  $RMSE_{N_e}$  rose again, in some cases sharply (Fig. S5, bottom, Supporting information).

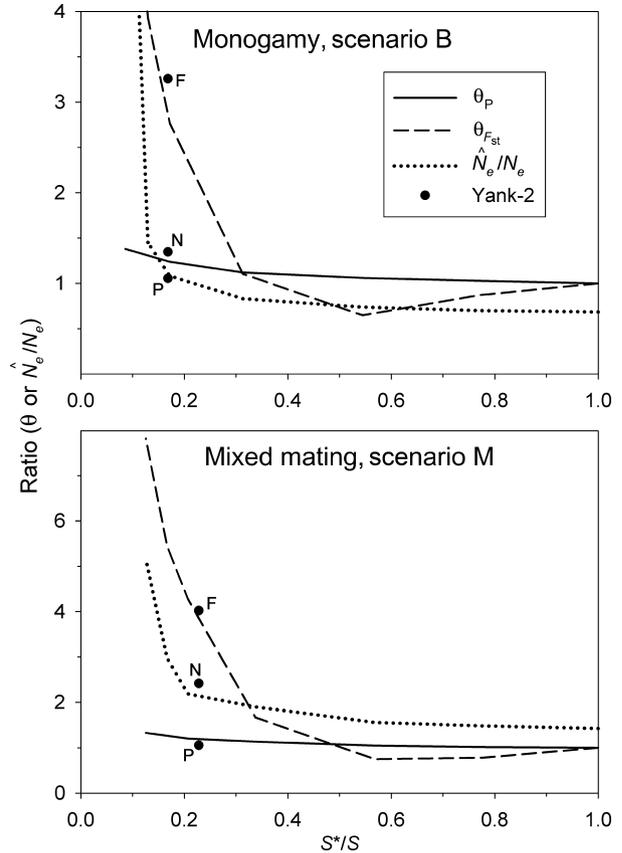
#### *Empirical example comparing methods for estimating allele frequency*

Under the assumption that the inferred pedigree was correct (‘Sample Related’ scenario), application of the BLUE maximized ESS for each of the 70 collections of juvenile coho salmon (Fig. 7, top; Supporting information). The optimal-z sibling removal strategy never decreased ESS and, in some collections, substantially increased effective sample size. ESS for the Yank-1 strategy was sometimes smaller than, but more often larger than, effective sample



**Fig. 7** Effective sample size (ESS) for various sibling removal and weighting strategies applied to empirical data for 70 collections of juvenile coho salmon. Collections are arranged along the  $x$ -axis in increasing order of ESS under the naïve scenario, which equally weights all individuals. Other scenarios either weight individuals according to the inferred pedigree (BLUE) or remove siblings according to rules described in the text. Top panel: the sibling reconstruction by Colony, which identified many sibling groups in most samples, is assumed to be correct. Bottom panel: genotypes are scrambled, mimicking samples in which all individuals are unrelated. In all but the naïve method, sibling adjustments are made based on results from Colony, which inferred spurious siblings groups in all samples. However, in only one collection (#61) was a family group with three or more full-siblings inferred, so no siblings were removed for the other collections under either the Yank-1 or Yank-2 methods, and ESS in those cases was identical to that of the naïve estimator.

size of the naïve estimator. For almost all collections, the Yank-2 strategy produced a higher ESS than both the naïve estimator and the Yank-1 strategy. This indicates that if there is large variance in family size, as appears to be the case in these collections of juvenile coho salmon, informed sibling removal can potentially improve estimates of allele frequency. In agreement with these theoretical results, when we applied the BLUE to simulated data for the scenarios in Table 2, in every case ESS based on BLUE-weighted allele frequency was higher than ESS for the full sample (Fig. S6, Supporting information).



**Fig. 8** Representative results from applying the Yank-2 sibling removal method to simulated data. For monogamy (top panel) and mixed mating (bottom panel) scenarios as described in Table 2, the solid, dashed and dotted lines show (respectively) results for probabilistic sibling removal for estimation of allele frequency (relative  $\text{RMSE}_P = \theta_P$ ), population differentiation (relative  $\text{RMSE}_{F_{ST}} = \theta_{F_{ST}}$ ) and effective population size ( $\hat{N}_e/N_e$ ). Filled circles show results for Yank-2, and small letters ( $P$ ,  $F$ ,  $N$ ) indicate which data series each Yank-2 data point is related to. The  $x$ -axis shows how much sibling removal reduces the usable sample size ( $S^*$ ) compared to the full sample ( $S$ ). In general, compared to probabilistic sibling removal schemes that produce the same overall reduction in sample size (indexed by  $S^*/S$ ), the Yank-2 method (i) leads to greater precision in estimating  $P$  (lower  $\theta_P$ ), (ii) less precision in estimating  $F_{ST}$  (higher  $\theta_{F_{ST}}$ ) and (iii) higher estimates of effective size (higher  $\hat{N}_e/N_e$ ). For complete results for Yank-2, see Table S1 (Supporting information).

When genotypes of the juvenile samples were scrambled to mimic those of unrelated individuals, a very different result was obtained: except in the smallest collections, in which Colony correctly inferred everyone to be unrelated, using the BLUE or the optimal-z strategy always reduced ESS compared to that of the naïve (and in this case, correct) estimator (Fig. 7, bottom). Reductions in ESS occurred in samples for which Colony erred by identifying spurious relationships. Colony is much more likely to spuriously identify HS relationships or pairs of FS than large, FS

groups from data where no family structure exists. Across all 70 permuted coho salmon collections, Colony inferred no FS groups of size four or greater, only one FS group of size three and 193 pairs of FS (Table S2, Supporting information). The Yank-1 and Yank-2 are designed so that sibling groups of size  $< 3$  do not trigger any removals, and as a consequence, they are relatively resilient to pairs of spuriously inferred FS. Accordingly, even using the incorrectly inferred pedigree, ESS for these two sibling removal strategies was identical to that of the naïve estimator in almost every collection; in the one exception, removing siblings using both the Yank-1 and Yank-2 protocols slightly reduced ESS.

#### *Yank-2 with simulated data*

Because the Yank-2 strategy appeared to perform relatively well with the empirical coho salmon data sets, we evaluated its performance with simulated data for the monogamy and mixed mating models (the random mating model produces few FS so we did not evaluate Yank-2 for this model). Full results are shown in Table S1 (Supporting information); representative patterns are illustrated in Fig. 8. In comparison with probabilistic sibling removal schemes that produce the same overall reduction in sample size (indexed by the ratio  $S^*/S$ ), Yank-2 generally increased precision of  $\hat{P}$  (and hence also increases ESS) but reduced precision of  $\hat{F}_{ST}$ . The Yank-2 equalizes family contributions to no more than two FS, so it also tends to increase  $\hat{N}_e$ . If sampling is random (as in the scenarios shown in Fig. 8), this leads to more upward bias in  $\hat{N}_e$ , while for nonrandom samples, it can reduce the downward bias in estimates of effective size.

## Discussion

### *Allele frequency*

Presence of family structure in a sample does not lead to systematic bias in estimation of parental allele frequency, in the sense that large numbers of siblings do not consistently lead to under- or overestimation of *TrueP*. However, for any given sample, the presence of siblings will tend to skew  $\hat{P}$  towards frequencies of those parents responsible for the large families, and that increases RMSE of, and reduces precision of,  $\hat{P}$ . The rationale for removing siblings to estimate allele frequency is twofold: (i) reducing large families to one or a small number of representatives produces a more balanced picture of allele frequencies in all the parents, and (ii) siblings provide partially redundant information about parental allele frequencies, especially when family size is large, so removing them does not sacrifice much useful information.

Our simulation results provide two general insights on this issue. First, in every combination of mating model and sampling strategy that we evaluated, removing some or all siblings increased the ratio of  $ESS_{\text{reduced}}$  to  $S^*$ . If all

individuals provided completely independent information about allele frequency, this ratio should not change with reductions in  $S^*$ . Therefore, this result confirms the partial redundancy of siblings with respect to estimation of *TrueP*. The second key result is that in most of the simulated scenarios, this partial redundancy was not strong enough to fully offset the loss of precision associated with reducing overall sample size. When sibling removal had any appreciable effect on performance, it always reduced precision of  $\hat{P}$ . In the simulations, the only exception to this pattern was for Scenario R, which led to greatly overdispersed variance in reproductive success among the parents. This suggests that purging putative siblings is not a universally sound strategy to improve estimates of *TrueP*; it can easily make things worse.

The empirical example using collections of juvenile coho salmon provides important context for interpreting results of the simulated populations. The true pedigree of these samples is not known, but based on the biology of the species, the nature of the samples, and the estimated sibships from Colony, the samples contain a great deal of family structure. We showed that, under the assumption that the inferred pedigree is correct, noninteger weighting of individuals using the BLUE approach produces an estimate of parental allele frequency with lower variance than the naïve estimate or any method that removes entire individuals. On the other hand, our results also show that performance of the BLUE method can be worse than the naïve estimator when the inferred pedigree is not correct. Although it might be tempting to apply the BLUE or try to find the optimal removal scheme given the inferred pedigree of sample members, there are clear risks associated with that approach.

Two variations of sibling removal (Yank-1 and Yank-2) that have been used in some empirical studies performed well in analysis of the coho salmon samples: in most cases, they produced a higher ESS than the naïve estimator when the inferred pedigree was assumed to be correct, and in all collections but one they did not degrade the performance of the allele frequency estimate when the pedigree included spurious family structure. The Yank-2 procedure consistently had a higher effective sample size than Yank-1, suggesting that it might be a generally better option. This empirical coho salmon example provides a somewhat more optimistic picture of the effectiveness of sibling removal for estimating allele frequency than did the simulations. We verified this general pattern by incorporating the Yank-2 procedure into the simulated data (Fig. 8, Table S1, Supporting information). We expect this result can be explained by two factors. First, the Yank-1 and Yank-2 methods only reduce families having three or more FS, whereas one of a pair of siblings could be removed during proportional sibling removal in the simulations. Second, in many of the coho salmon samples, Colony inferred not only some large FS families but also many unrelated individuals (see Table S2, Supporting information)—a feature that also characterized the one scenario in the simulations (Scenario R) for which aggressive purging

of siblings did substantially reduce  $RMSE_p$ . Therefore, results for the empirical example support the premise from the simulations that removal of putative siblings is most likely to improve allele frequency estimation when the samples include a substantial degree of family structure and a large variance in family size. If every family is large, there will be little to gain by downsampling some of them. In the simulations, sibling removal degraded performance most in scenarios (e.g. A, B, S, U; Figs 3 and 4) where sample size was large compared to  $N_e$ , which tends to produce many large families. However, if there are only a few large families and many unrelated individuals, then the relative contribution of the large families to the estimated allele frequency can be brought into line by downsampling them (as in Scenario R). Plotting the distribution of inferred family sizes in empirical samples (see, for example, Whiteley *et al.* 2015) can be useful to researchers in this regard.

A caveat worth noting regarding the empirical example is that only one type of sibling misidentification was modelled (spurious identification of related individuals when all individuals are unrelated). The Yank-1 and Yank-2 approaches, although they are immune to spuriously inferred HS and FS groups larger than 3, could perform worse if FS groups larger than 3 were inferred. Given the genotype data that we had (95 SNPs), Colony almost never identifies such large sibling groups among totally unrelated individuals; however, it is possible that it could spuriously identify large FS groups among samples with many HS, or if fewer loci are used in the analysis. Thus, practitioners should be aware that our results do not provide information about how the Yank-1 and Yank-2 methods might perform in their own data sets when large family groups are inferred by unreliable pedigree reconstruction methods. These should be treated on a case-by-case basis. Tools for evaluating expected performance of sibling removal methods are available at <https://github.com/eriqande/afblue>.

### Population differentiation

Our results for  $F_{ST}$  estimation largely paralleled those for allele frequency, which suggests that they might be generally applicable to other types of analyses that strongly depend on population allele frequencies, such as assignment tests and genetic mixture analysis—but that is only a conjecture that requires empirical evaluation. We did find that a wider range of scenarios showed improved performance of  $\hat{F}_{ST}$  under partial removal of siblings (see Figs 3 and 4). We suspect this is because removing siblings reduces the upward bias in  $\hat{F}_{ST}$  that occurs when family structure inflates the variance of  $\hat{P}$ . Although family structure does not bias  $\hat{P}$  for a single population,  $\hat{F}_{ST}$  is proportional to the variance of  $\hat{P}$  across populations, and that is affected by relatedness within samples.

Because using the BLUE increases precision of  $\hat{P}$  when the inferred pedigree is correct, it seems likely that it also should improve performance of  $\hat{F}_{ST}$  under those conditions. However, our results also suggest that using BLUE could also degrade performance of  $\hat{F}_{ST}$  when the pedigree

is not reliable. Furthermore, different estimators of  $F_{ST}$  and related quantities have different sensitivities to effects of rare alleles, the number of populations exchanging genes and corrections for sampling error (Bhatia *et al.* 2013). Finally, in our simulations, the Yank-2 method generally reduced precision in  $\hat{F}_{ST}$ , even though it generally increased effective sample size and precision in  $\hat{P}$ . Therefore, more empirical evaluations are needed before any general conclusions can be drawn about optimal ways of dealing with siblings to estimate measures of genetic differentiation.

### Effective population size

Presence of siblings affects estimation of  $N_e$  differently (in two ways) than it does estimation of allele frequency. First, whereas family structure only affects precision of  $\hat{P}$ , it can severely bias estimates of  $N_e$ . Second, whether samples are collected randomly or not has a much larger effect on  $\hat{N}_e$  than  $\hat{P}$ . For the LD and sibship methods, family structure that occurs naturally in all finite populations is part of the signal that allows one to estimate effective size. Nonrandom sampling that tends to collect groups of siblings together creates additional family structure that is mistaken for drift, which downwardly biases  $\hat{N}_e$ . In theory, an unbiased sample could be reconstructed by removing siblings collected in excess of the random expectation (Fig. 6). In practice, however, this will be difficult or impossible to accomplish. It is clear even from our limited evaluations that no generic, one-size-fits-all strategy of fractional sibling removal will be effective in every scenario. The location of the sweet spot of sibling removal that produces a completely unbiased estimate varies widely, depending on (at least) the mating system, type of sampling, values of  $S$  and  $N_e$  and whether all siblings or only FS are removed. Furthermore, the Goldilocks zone (an area around the sweet spot that represents ‘just the right amount’ of bias adjustment to produce a reasonable estimate) can be very narrow, such that the consequences of small errors in identifying the optimal fraction of siblings to remove can be harsh. For example, under Scenario W, an unbiased estimate can be achieved by removing about 95% of all siblings, but removing only 75% leaves  $\hat{N}_e$  downwardly biased by 70%, while removing all siblings produces an estimate that is over five times as large as true  $N_e$  (Fig. 6). Researchers interested in pursuing this option are faced with a chicken-and-egg conundrum: only if one knows the true effective population size can one determine the precise amount of sibling reduction that will produce an unbiased estimate of  $N_e$ .

We used the LD method to estimate  $N_e$ , but similar results can be expected for Wang’s (2009) sibship method and for ONeSAMP (Tallmon *et al.* 2008); the latter’s approximate-Bayesian-computation program uses several summary statistics, but the most important signal is from LD. In the standard temporal method for estimating  $N_e$ , temporal  $F$  is a function of the variance of  $\hat{P}$  in the two time periods (Nei & Tajima 1981; Waples 1989); therefore,

anything that affects precision of  $\hat{P}$  can affect both bias and precision of  $\hat{F}$ , just as it can for  $\hat{F}_{ST}$ . This topic merits some empirical evaluation as well.

#### *Random and nonrandom sampling*

We see no realistic way to distinguish random and nonrandom samples based on patterns of relatedness, even if family structure can be reconstructed with 100% accuracy. For any given array of sibling relationships that can be generated with nonrandom sampling, it is possible to imagine a random sample that could produce the same result. For example, consider Scenario R, in which most of the  $N$  potential parents produced no offspring that appeared in the samples, while one or a few lucky pairs produced large numbers of offspring. This is exactly the kind of result that could arise if offspring of a species with Type III survival are sampled at an early age before they have properly mixed (as could easily occur with many fish, marine invertebrates, amphibians, insects, etc.). On the other hand, this pattern of family structure (realized  $N_e$  was <2% of  $N$  for Scenario R; Table 2) also conforms to predictions of Hedgecock's (1994) hypothesis of sweepstakes reproductive success (reviewed by Hedgecock & Pudovkin 2011), which has been postulated to be responsible for a number of tiny estimates of the  $N_e/N$  ratio in marine species (Hedrick 2005; Hauser & Carvalho 2008; Waples 2016). A sample with this type of family structure could therefore represent an extreme case of nonrandom sampling, or an extreme case of naturally overdispersed variance in reproductive success that characterizes a novel and important evolutionary phenomenon. Without independent information about the nature of the samples, it generally will not be possible to distinguish between these two possibilities.

#### **Conclusions and future directions**

Identifying optimal ways for dealing with siblings in population genetic data sets is a complex problem with no single, one-size-fits-all solution. Our evaluations of simulated and empirical data sets by no means represent a comprehensive evaluation of this topic, nor was that our intent. Nevertheless, several important points can be made.

1 The paradigm that seems to be emerging in applied conservation genetics is that family structure in genetic data sets is an aberration that has to be cured by removing putative siblings. We think it is dangerous to adopt this as a general perspective, even though it might be generally effective in some specific applications, such as Bayesian clustering, and can be conditionally effective in some other applications, such as estimation of allele frequency. However, even our limited evaluations demonstrate that the consequences of adjusting for putative siblings can depend on the nature of the downstream analyses, the method for adjusting for siblings, the reliability of the inferred pedigree, and the magnitude and distribution of family structure in the sample, among

many other potentially contributing factors. Until sufficient empirical evaluations have been conducted that closely model these factors in a way that is relevant for a specific intended application, it would be prudent for researchers to take some guidance from the Hippocratic Oath: 'first, do no harm'. That is, the first objective should be to ensure that any data manipulations to account for putative siblings do not make problems associated with family structure worse. Our results indicate that is a likely consequence of indiscriminate purging of putative siblings.

- 2 Removing putative siblings, or downweighting them using an approach like BLUE, is not a panacea for dealing with problems arising from nonrandom sampling. Even if sibship reconstruction is 100% accurate and an optimal adjustment for allele frequency can be made, the result will still be lower precision than could have been obtained with a random sample. And for analyses such as  $N_e$  estimation that can be biased by excess family structure in nonrandom samples, it will be difficult or impossible to determine the optimal fraction of siblings to remove. Therefore, if one strongly suspects they have nonrandom, family-correlated samples, a far better strategy is to go back and obtain a random sample. Of course, that is easier said than done in many natural populations.
- 3 The simulation results were rather pessimistic regarding removal of putative siblings, except with very strong family structure. Nevertheless, those results are optimistic because they all assumed that every true FS and HS could be correctly identified, and that no unrelated pairs were mistakenly identified as siblings. Although that level of precision might be approached under some ideal conditions, in general there will be uncertainties associated with many sibling assignments. The empirical example using samples of coho salmon demonstrated that errors in pedigree reconstruction can degrade performance of sibling adjustments. Therefore, a rigorous assessment of performance of sibling adjustment methods must include realistic assumptions about the reliability of sibling reconstruction.
- 4 Considerably more empirical and theoretical evaluations are needed of this complex problem. Other mating and sampling systems could be modelled, using a wider range of parameter values, and a wider range of downstream analyses could be evaluated. We ignored complications such as age structure and migration, which are important for many natural systems. The following questions seem relevant: Can classes of analytical methods be identified that respond in predictable ways? If different analyses show different sensitivities to family structure, then a likely consequence would be that different methods use different subsets of the original data. How can these differences be accommodated in trying to integrate results to draw population-level inferences about key evolutionary and ecological processes?
- 5 Finally, our analyses related to allele frequency assumed that the goal is to estimate parental allele frequency,

defined to be  $P$  in the entire population of  $N$  potential parents. But what if instead one wants to estimate  $P$  in the same generation that is sampled? Or what if one wants to estimate the  $P$  that characterizes the effective population that actually produces the next generation? In the latter case, the best strategy would be to take a large, random sample of progeny and weight all individuals equally, regardless of family structure. So different perspectives about what quantity we want to estimate can lead to different conclusions regarding handling of siblings. Researchers should give this some thought before considering various sibling adjustment options.

### Acknowledgements

We appreciate useful comments by and discussions with Mike Ford, Rus Hoelzel, Daniel Ruzzante, Jinliang Wang, Ryan Waples, Andrew Whiteley and two anonymous reviewers. Genetic data for the coho salmon samples were kindly provided by John Carlos Garza and Libby Gilbert-Horvath from the Southwest Fisheries Science Center.

### References

- Ackerman MW, Hand BK, Waples RK *et al.* (2017) Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary Applications*, **10**, 146–160.
- Almudevar A, Anderson EC (2012) A new version of PRT software for sibling groups reconstruction with comments regarding several issues in the sibling reconstruction problem. *Molecular Ecology Resources*, **12**, 164–178.
- Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. *Molecular Ecology Resources*, **8**, 1219–1229.
- Ashley MV, Caballero IC, Chaovalitwongse W *et al.* (2009) KINALYZER, a computer program for reconstructing sibling groups. *Molecular Ecology Resources*, **9**, 1127–1131.
- Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution*, **58**, 1891–1900.
- Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting  $F_{ST}$ : the impact of rare variants. *Genome Research*, **23**, 1514–1521.
- Chakraborty R, Leimar O (1987) Genetic variation within a subdivided population. In: *Population Genetics and Fishery Management* (eds Ryman N, Utter F), pp. 89–102. University of Washington Press, Seattle, Washington.
- Garza JC, Gilbert-Horvath EA, Spence BC *et al.* (2014) Population structure of steelhead in coastal California. *Transactions of the American Fisheries Society*, **143**, 134–152.
- Gilbert-Horvath EA, Pupal KA, Spence BC, Williams TH, Garza JC (2016) Hierarchical phylogeographic structure of coho salmon in California. *Transactions of the American Fisheries Society*, **145**, 1122–1138.
- Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources*, **10**, 304–313.
- Hansen MM, Nielsen EE, Mensberg KLD (1997) The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout *Salmo trutta* L. *Molecular Ecology*, **6**, 469–474.
- Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.
- Hedgcock D (1994) Does variance in reproductive success limit effective population size of marine organisms? In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont A), pp. 122–134. Chapman & Hall, London.
- Hedgcock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, **87**, 971–1002.
- Hedrick P (2005) Large variance in reproductive success and the  $N_e/N$  ratio. *Evolution*, **59**, 1596–1599.
- Hess JE, Campbell NR, Docker MF *et al.* (2015) Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, **15**, 187–202.
- Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–555.
- Linlökken AN, Haugen TO, Mathew PK, Johansen W, Lien S (2016) Comparing estimates of number of breeders  $N_b$  based on microsatellites and single nucleotide polymorphism of three groups of brown trout (*Salmo trutta* L.). *Fisheries Management and Ecology*, **23**, 152–160.
- McPeck MS, Wu X, Ober C (2004) Best linear unbiased allele-frequency estimation in complex pedigrees. *Biometrics*, **60**, 359–367.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA*, **70**, 3321–3323.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics*, **98**, 625–640.
- Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. *PeerJ*, **4**, e1813.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- R Core Team (RCT) (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.
- Ringler E, Mangione R, Ringler M (2015) Where have all the tadpoles gone? Individual genetic tracking of amphibian larvae until adulthood. *Molecular Ecology Resources*, **15**, 737–746.
- Rodríguez-Ramilo ST, Toro MA, Wang J, Fernandez J (2014) Improving the inference of population genetic structure in the presence of related individuals. *Genetics Research*, **96**, e003.
- Rodríguez-Ramilo ST, Wang J (2012) The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources*, **12**, 873–884.
- Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) OneSamp: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, **8**, 299–301.
- Wang J (2009) A new method for estimating effective population size from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148–2164.
- Wang J, Santure AW (2009) Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, **181**, 579–1594.
- Wang J (2004) Sibship reconstruction from genetic data with typing errors. *Genetics*, **166**, 1963–1979.

- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Waples RS (2016) Tiny estimates of the  $N_e/N$  ratio in marine fishes: are they real? *Journal of Fish Biology*, **89**, 2479–2504.
- Waples RS, Do C (2008) *LDNe*: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753–756.
- Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**, 244–262.
- Waples RS, Waples RK (2011) Inbreeding effective population size and parentage analysis without parents. *Molecular Ecology Resources*, **11**(Suppl. 1), 162–171.
- Whiteley AR, Coombs JA, Hudy M, Robinson Z, Nislow KH, Letcher BH (2012) Sampling strategies for estimating brook trout effective population size. *Conservation Genetics*, **13**, 625–637.
- Whiteley AR, Coombs JA, Cembrola M *et al.* (2015) Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. *Molecular Ecology*, **24**, 3585–3602.
- Wright S (1938) Size of population and breeding structure in relation to evolution. *Science*, **87**, 430–431.
- Wright S (1978) *Evolution and the Genetics of Populations. Vol. IV. Variability Within and Among Natural Populations*. University of Chicago Press, Chicago, Illinois.

---

R.W. conceived the study. R.W. and E.A. conducted the analyses and wrote and edited the manuscript.

---

doi: 10.1111/mec.14022

### Data accessibility

Analyses of effective sample size for the coho salmon collections were carried out using an R package called `AFBLUE`,

available at <https://github.com/eriqande/afblue>, which includes code and data to replicate all reported results. R code for the simulations and related analyses is also available at this site. Full-sibling families inferred by Colony for the coho salmon collections are listed in Table S2 (Supporting information).

### Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Simulation results.

**Table S2** Analyses of collections of juvenile coho salmon.

**Fig. S1** Relationship between observed and expected fractions of siblings in simulated populations.

**Fig. S2** Proportions of all pairwise relationships that are full siblings and half siblings, for three different mating models and random and non-random samples.

**Fig. S3** Relationship between true  $N_e$  and  $N_e$  estimated from the parentage-analysis-without-parents (PWOP) method (eqn 3, main text).

**Fig. S4** Relationship between  $\hat{N}_e$  estimated from LDNe and realized  $N_e$  calculated using PWOP (eqn 3), for two scenarios using random sampling and the monogamy mating model.

**Fig. S5** Relative RMSE of  $1/\hat{N}_e$  for random and non-random samples under three mating models.

**Fig. S6** Performance of the BLUE method with simulated data.

**Appendix S1** Detailed methods.